# Numerical Approximation
# of a Wave Equation with Unilateral Constraints

## By Michelle Schatzman and Michel Bercovier

**Abstract.** The system $u_{tt} - u_{xx} \ni f$, $x \in (0, L) \times (0, T)$, with initial data $u(x, 0) = u_0(x)$, $u_t(x, 0) = u_1(x)$ almost everywhere on $(0, L)$ and boundary conditions $u(0, t) = 0$, for all $t \geq 0$, and the unilateral condition

$$u_x(L, t) \geq 0, u(L, t) \geq k_0, (u(L, t) - k_0)u_x(L, t) = 0$$

models the longitudinal vibrations of a rod, whose motion is limited by a rigid obstacle at one end. A new variational formulation is given; existence and uniqueness are proved. Finite elements and finite difference schemes are given, and their convergence is proved. Numerical experiments are reported; the characteristic schemes perform better in terms of accuracy, and the subcharacteristic schemes look better.

**1. Introduction.** Consider the following problem: a linear rod vibrates longitudinally; one end of the rod is fixed, and the other one is free to move, as long as it does not hit a material obstacle. This obstacle may constrain the displacement of this extremity either to be greater than or equal to some given number, or to be smaller than or equal to some number.
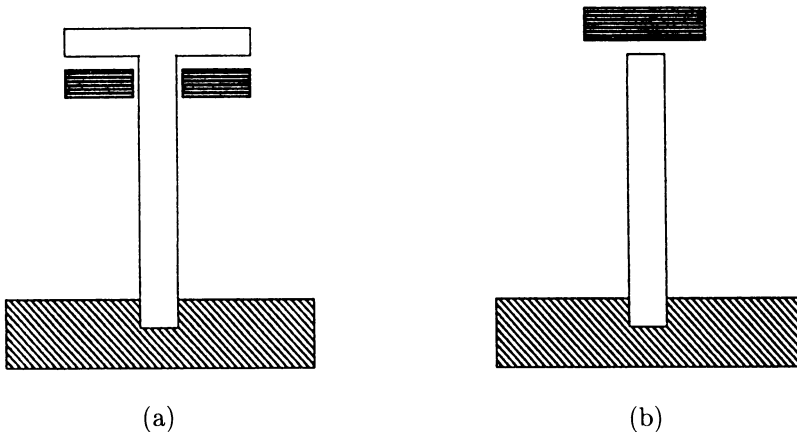


FIGURE 1
(a) lower constraint: $u(L, t) \geq k_0$; (b) upper constraint: $u(L, t) \leq k_0$.

We can describe this situation mathematically as follows: assume that the material of the rod is elastic, homogeneous, linear, and make the approximation of small displacements. Let $x$ be the spatial coordinate along the rod, with the origin at the fixed end; let $u(x,t)$ be the displacement at time $t$ of the material point of spatial coordinate $x$ at rest. Let $f$ denote a density of exterior forces, depending on space and time. With an adequate scaling, the velocity of waves in the rod is one and the length of the rod at rest is $L$. The displacement $u$ satisfies the following equation:

$$(1.1) \qquad \Box u \equiv u_{tt} - u_{xx} = f \quad \text{in } Q_T \equiv (0, L) \times (0, T).$$

The boundary conditions are detemined as follows: at the fixed end we have a Dirichlet boundary condition

$$(1.2) \qquad u(0, t) = 0.$$

To be definite, assume a lower obstacle at the other end; then

$$(1.3) \qquad u(L, t) \geq k_0.$$

When the rod touches the obstacle, its reaction can be only upwards (see Figure 1), so that

$$u_x(L, t) \geq 0 \quad \text{on the set } \{t \, / \, u(L, t) = k_0\};$$

when the rod does not touch the obstacle, the end is free

$$u_x(L, t) = 0 \quad \text{on the set } \{t \, / \, u(L, t) > k_0\}.$$

These last two conditions, which will be entirely justified in the sequel, can be summarized as

$$(1.4) \qquad u_x(L, t) \geq 0,$$

$$(1.5) \qquad u_x(L, t)(u(L, t) - k_0) = 0.$$

Conditions (1.3)–(1.5) are usually termed "unilateral conditions". We are given initial conditions

$$(1.6) \qquad u(x, 0) = u_0(x),$$

$$(1.7) \qquad u_t(x, 0) = u_1(x).$$

It has been shown in [6, Theorem 14] that, if $u_0$ belongs to the Sobolev space $H^{3/2}(0, L)$, $u_0(0) = 0$, $u_0(L) \geq k_0$, if $u_1$ belongs to the Sobolev space $H^{1/2}(0, L)$, and if $f$ belongs to $H^{3/2}(Q_T)$, then (1.1)–(1.7) possesses a unique solution $u$ in the space

$$L^\infty(0, T; H^{3/2}(0, L)) \cap W^{1,\infty}(0, T; H^{1/2}(0, L)).$$

It has been proved in [6] that the energy of the solution is conserved:

$$
\begin{aligned}
(1.8) \qquad &\frac{1}{2} \int_0^L \left( |u_t|^2 + |u_x|^2 \right)(x, t)\, dx \\
&= \frac{1}{2} \int_0^L \left( |u_1|^2 + \left| \frac{du_0}{dx} \right|^2 \right)(x)\, dx \\
&\quad + \int_0^t \int_0^L u_t(x, s)\, f(x, s)\, dx\, ds, \quad \forall t \in [0, T].
\end{aligned}
$$

Relation (1.8) is a consequence of the equations, and therefore, the model considered here does not include the possibility of a loss of energy during the contact with the obstacle.

The problem described here is closely related to the problem of a vibrating string with pointwise unilateral constraints which has been studied in [1], [3], [4], [11] and [12]. Let $u$ be a solution of (1.1)–(1.7), and denote by $\tilde{u}$ the function defined by

$$\tilde{u}(x,t) = u(x,t), \qquad 0 \leq x \leq L,$$
$$\tilde{u}(x,t) = u(2L - x,t), \qquad L \leq x \leq 2L;$$

the initial conditions are symmetrized similarly. Then it is easy to check that $\tilde{u}$ satisfies the following relations, where $\mu$ is a measure,

(1.9) $$\Box \tilde{u} = \mu \quad \text{over } (0,2L) \times (0,T),$$

(1.10) $$\tilde{u}(L,t) \geq k_0 \quad \text{for } t \text{ in } (0,T),$$

(1.11) $$\operatorname{supp}(\mu) \subset \{L\} \times \{t \in [0,T] \,/\, \tilde{u}(L,t) = k_0\},$$

(1.12) $$\tilde{u}(x,0) = \tilde{u}_0(x),$$

(1.13) $$\tilde{u}_t(x,0) = \tilde{u}_1(x),$$

(1.14) $$\tilde{u}(0,t) = \tilde{u}(2L,t) = 0.$$

Only relation (1.9) is not completely obvious: $\Box \tilde{u}$ is supported in $\{L\} \times [0,T]$; $\Box \tilde{u}$ is not zero whenever $\tilde{u}_x$ jumps; but only negative jumps are allowed, thanks to condition (1.4) and to the symmetry of $\tilde{u}$. With the minus sign before the second derivative in space, this justifies intuitively (1.9). The argument can be made rigorous with standard distribution techniques.

Problem (1.9)–(1.14) has been studied in [11], where a solution is obtained in larger spaces of initial data than in [6]. Reference [11] gave a variational formulation of (1.9)–(1.14), which was rather awkward. Problem (1.1)–(1.7) possesses an explicit solution (see [6]), because the medium is uniform. A new variational formulation is given here; it has two aims: the first is to give a good framework in which to generalize the present homogeneous problem to a nonhomogeneous one, and the second to develop numerical schemes in a natural fashion.

In Section 2 of this paper, we give our variational formulation and sketch the proof of the equivalence of the variational formulation with the original formulation.

In Section 3 we state an existence theorem for a solution by Galerkin's method of approximation in finite-dimensional space and outline its proof . In the remainder of this section, we define a family of numerical schemes with the help of a variational formulation. Let $V_h$ be a sequence of approximation spaces of

(1.15) $$V \equiv \{u \in H^1(0,L) \,/\, u(0) = 0\};$$

$V_h$ can be a space of finite elements; the convex set $K_h$ is the set of elements of $V_h$ which satisfy the constraint; let $(\cdot,\cdot)$ denote the scalar product in $L^2(0,L)$, and let $a$ denote the scalar product in $V$ defined by

(1.16) $$a(u,v) = \int_0^L u_x \, v_x \, dx.$$

Then, our family of schemes is defined by

$$
(1.17) \qquad
\begin{cases}
u^{n+1} \in K_h; \\[2mm]
\left( \dfrac{u^{n+1} - 2u^n + u^{n-1}}{\Delta t^2}, v - u \right) + a(u^n, v - u^{n+1}) \\[4mm]
\qquad \geq (f^n, v - u^{n+1}), \quad \forall v \in K_h.
\end{cases}
$$

Here, $f^n$ will be a suitable approximation of $f(\cdot, n\Delta t)$, and the initial data $u^0$ and $u^1$ of the scheme are adequately chosen.

The scheme (1.17) converges under a stability condition which is the same as the condition of the linear case.

In Section 4 we turn to explicit difference schemes; they can be written in variational form and are analogous to (1.17), with the main difference that the mass matrix is diagonal, so that their implementation is very easy. These schemes are given by

$$
(1.18) \qquad
\begin{aligned}
& \frac{U_j^{k+1} - 2U_j^k + U_j^{k-1}}{\Delta t^2} - \frac{U_{j+1}^k - 2U_j^k + U_{j-1}^k}{\Delta x^2} = F_j^k \quad \text{if } j < n, \\[3mm]
& U_n^{k+1} = \max\left( k_0, 2U_n^k - U_n^{k-1} + \frac{\Delta t^2}{\Delta x^2}\left( U_{n-1}^k - U_n^k \right) + \Delta t^2 F_n^k \right).
\end{aligned}
$$

The proof of convergence is fairly easy if the scheme is noncharacteristic ($\Delta t < \Delta x$); it is much more technical in the characteristic case ($\Delta t = \Delta x$).

In Section 5 we report on numerical experiments. They can be summarized as follows: if the initial data are smooth, the characteristic and noncharacteristic schemes give reasonably good results; the characteristic scheme is substantially better. This suggests that for nonhomogeneous rods, the CFL number should be taken as large as possible. If the initial data are piecewise affine, the characteristic scheme gives good results, while the noncharacteristic scheme gives very bad results. Thus dispersion does not seem to make matters easier.

We surmise that the computation of phenomena with rough data in a nonhomogeneous medium will be difficult.

## 2. The Variational Formulation.

2.1. *Notations and Definitions.* We shall need a number of spaces and sets. The first of these is $V$, which has already been defined by (1.15), with scalar product $a$ defined by (1.16); the norm on $V$ is denoted $\|\cdot\|$. The space $L^2(0, L)$ is alternatively denoted $H$, with the standard scalar product denoted $(\cdot, \cdot)$, and associated norm denoted $|\cdot|$. The convex set $K$ is defined as

$$
(2.1) \qquad K = \{ u \in V \, / \, u(L) \geq k_0 \}.
$$

We define now spaces involving the time variable:

$$
(2.2) \qquad \mathscr{X}_2 = \{ u \in L^2(0, T; V) \, / \, u_t \in L^2(0, T; H) \},
$$

with norm

$$
(2.3) \qquad [u]_2 = \left[ \int_0^T \left( \|u(t)\|^2 + |u_t(t)|^2 \right) dt \right]^{1/2};
$$

similarly, we define

$$\text{(2.4)} \qquad \mathscr{X}_\infty = \{u \in L^\infty(0,T;V) \,/\, u_t \in L^\infty(0,T;H)\},$$

with norm

$$\text{(2.5)} \qquad [\![u]\!]_\infty = \operatorname*{ess\,sup}_{t \in [0,T]}(\|u(t)\| + |u_t(t)|).$$

Observe that $\mathscr{X}_2$ is the space of functions of locally bounded energy; its topology is the topology of $H^1(Q_T)$. In particular, the elements of $\mathscr{X}_2$ have a trace on $\{x = L\}$. Therefore, the following definition makes sense:

$$\text{(2.6)} \qquad \mathscr{K} = \{u \in \mathscr{X}_2 \,/\, u(L,t) \geq k_0 \text{ almost everywhere}\},$$

and it is equivalent to

$$\text{(2.7)} \qquad \mathscr{K} = \{u \in \mathscr{X}_2 \,/\, u(\cdot,t) \in K \text{ for almost every } t\}.$$

2.2. *The Variational Formulation.* The main result of this section is the following theorem of equivalence:

THEOREM 2.1. *Let $f$ belong to $L^2(Q_T)$, let $u_0$ belong to $K$ and $u_1$ to $H$. Let $u$ belong to $\mathscr{X}_2$. Then $u$ is a solution of* (1.1)–(1.7) *if and only if*

$$\text{(2.8)} \qquad u \in \mathscr{K},$$

$$\text{(2.9)} \qquad u(0) = u_0,$$

$$\text{(2.10)} \qquad -(u_1, v(0)) - \int_0^T (u_t, v_t - u_t)\,dt + \int_0^T a(u, v - u)\,dt \geq \int_0^T (f, v - u)\,dt,$$

$$\forall v \in \mathscr{K} \text{ such that there exists } \eta > 0 \text{ such that } v = u \text{ for } t \geq T - \eta.$$

At a naïve level, the equivalence between (2.8)–(2.10) and the original equations (1.1)–(1.7) is only a matter of writing enough Green formulae. The difficulty lies in the validation of the formal computations, and in particular in the trace theorems. It is not obvious a priori that $u_x$ has a trace on $\{L\} \times [0,T]$ because we require $u$ to be only in a space of functions of bounded energy, namely $\mathscr{X}_2$. Microlocal techniques could be used to prove that this trace exists. An alternative method is proposed in [12], with ample details. We give here only the main steps.

The existence of a trace for derivatives will depend on results relative to strongly continuous semigroups and their duals: let $X$ be a reflexive Banach space with norm $\|\ \|$, and dual $X^*$. Let $A$ be an operator from $D(A) \subset X$ into $X$. Assume that $A$ is the generator of a strongly continuous semigroup $S(t)$. We know that for every $u_0$ in $D(A)$ there exists a unique function $u$ in $C^1([0,T];D(A)) \cap C^0([0,T];X)$ such that

$$\text{(2.11)} \qquad \frac{du}{dt}(t) + Au(t) = 0$$

with initial condition

$$\text{(2.12)} \qquad u(0) = u_0.$$

Such a function is called a strong solution. As $X$ is reflexive, it is possible to define [8] nicely a dual semigroup $S^*(t)$ of $S(t)$; $S^*(t)$ is a strongly continuous semigroup of generator $A^*$, with

$$\langle u, A^* u^* \rangle = \langle Au, u^* \rangle, \quad \forall u \in D(A), \quad \forall u^* \in D(A^*).$$

The dual semigroup enables us to define a notion of weak solution of (2.11)–(2.12); for this purpose, we need a notation: the space $L^p_{\mathrm{loc}}((a,b); Z)$ is the space of measurable functions from $(a,b)$ to the Banach space $Z$, such that their restrictions to the compact subsets $[c,d]$ are in $L^p(c,d)$.

DEFINITION 2.2. *Let $u$ belong to $L^1_{\mathrm{loc}}((0,T); X)$ and $f$ to $L^1(0,T; X)$. Then $u$ is said to be a weak solution of* (2.11) *if and only if*

$$(2.13) \quad \begin{aligned} -\int_0^T \dot{\phi}(s)\langle u(s), u^* \rangle \, ds + \int_0^T \phi(s)\langle u(s), A^* u^* \rangle \, ds \\ = \int_0^T \phi(s)\langle f(s), u^* \rangle ds, \quad \forall \phi \in \mathscr{D}(0,T), \forall u^* \in D(A^*). \end{aligned}$$

Then we have the following regularity result on weak solutions.

PROPOSITION 2.3. *Let $u$ be a weak solution of* (2.11), *according to Definition 2.2. Then there exists a function $\tilde{u}$, equivalent to $u$ modulo null functions, such that*

$$\tilde{u} \in C^0([0,T]; X).$$

The proof of this result is given in [12].

It follows from Proposition 2.3 that a weak solution of (2.11) has an initial value $u(0)$. In particular, a weak solution of (2.11) is given by

$$u(t) = S(t)u_0.$$

A straightforward consequence of Proposition 2.3 is

COROLLARY 2.4. *Let $u$ belong to $\mathscr{X}_2$, and assume that $u_{tt} - u_{xx}$ (defined in the sense of distributions on $Q_T$) is square-integrable. Then, for every positive $\varepsilon$,*

$$(2.14) \qquad u \in C^0([0,T]; H^1(0, L-\varepsilon)) \quad and \quad u_t \in C^0([0,T]; L^2(0, L-\varepsilon)),$$

$$(2.15) \qquad u \in C^0([0,L]; H^1(\varepsilon, T-\varepsilon)) \quad and \quad u_t \in C^0([0,L]; L^2(\varepsilon, L-\varepsilon)).$$

In order to prove this result, we must choose a convenient semigroup; we have already a partial differential operator, $-\partial^2/\partial x^2$ for (2.14), and $-\partial^2/\partial t^2$ for (2.15). We only need boundary conditions. We choose Dirichlet boundary conditions. They are not satisfied by $u$, but by $u\phi$, where $\phi$ vanishes for $x = L$, for (2.14), and $\phi$ vanishes for $t = 0$ and $t = L$, for (2.15). In this argument we use the symmetry of $\Box$ with respect to time and space differentiation. Details can be found in [12].

A last technical lemma concerns $\mathscr{X}_\infty$ :

LEMMA 2.5. *The space $\mathscr{X}_\infty$ is included in the space of Hölder continuous functions $C^{0,1/2}(Q_T)$, and the injection from $\mathscr{X}_\infty$ to $C^{0,1/2}(Q_T)$ is continuous.*

The proof of this result is elementary. Details, if needed, can be found in [12].

The proof of Theorem 2.1 goes as follows: let $u$ satisfy (2.8)–(2.10); relation (2.8) makes sense because $u$ and $v$ belong to $\mathscr{X}_2$, the elements of which have traces on $[0, L] \times \{0\}$ belonging to $H^{1/2}(0, L)$. All the terms of (2.10) make sense.

It is immediate that $u$ satisfies

$$\Box u = f$$

in the sense of distributions.

Fix $\eta > 0$ and let $\mathscr{X}_2^\varepsilon$ be the space of functions belonging to $\mathscr{X}_2$ which vanish for $t \geq T - \eta$, and for $|x - L| + t \leq \varepsilon$. This space is tailored so as to avoid any (possible) difficulty with $u_t$ and $u_x$ in the corner $(0, L)$.

By integration by parts and density,

$$\int_0^T u_x(L, t)(v(L, t) - u(L, t))\, dt \geq 0, \quad \forall v \in \mathscr{K} \text{ such that } v - u \in \mathscr{X}_2^\varepsilon.$$

This makes sense, in view of Corollary 2.4. A classical convexity argument and a number of passages to the limit conclude this part of the proof.

Conversely, let $u$ satisfy (1.1)–(1.7). As $u$ belongs to $\mathscr{X}_2$ and $f$ to $L^2(Q_T)$, $u_x$ has a trace over $\{L\} \times [0, T]$, and (1.3)–(1.5) make sense. One has to check only (2.10), the other relations being clear. The main step is to show that $u_x(\cdot, t)$ belongs to $L^2(0, T - \eta)$ for all positive $\eta$; the difficulty is with the corner $(L, 0)$. Thus, we have a lemma:

LEMMA 2.6. *Let $u$ satisfy (1.1)–(1.7). Then for all positive $\eta$, $u$ belongs to $C^0([0, L]; H^1(0, T - \eta))$ and $u_t$ belongs to $C^0([0, L]; L^2(0, T - \eta))$.*

The proof of this result exploits the local energy identity inside $Q_T$, and Proposition 2.3. The variational inequality (2.10) holds for functions $v$ in $\mathscr{K}$ such that $u - v$ is in $\mathscr{X}_2^\varepsilon$. By density, and the usual limiting arguments, it holds for all functions $v$ in $\mathscr{K}$.

The proofs are presented in detail in [12], to which the reader is referred for further information. The uniqueness is proved in [6] and in [11], with different techniques.

## 3. Semidiscretization in Space and Finite Elements.

3.1. *Existence by Semidiscretization in Space.* Let $V_h$ be a sequence of finite-dimensional subspaces of $V$ such that

(3.1)
$$\overline{\bigcup_h V_h}^{\,V} = V.$$

The scalar product $(\cdot, \cdot)$ on $V_h$ is the restriction of the scalar product of $H$ to $V_h$.

The sequence of convex sets $K_h$ is defined by

(3.2)
$$K_h = K \cap V_h.$$

We shall denote by $M^1([0, T]; V_h)$ the space of bounded measures on $[0, T]$ with values in $V_h$. The space $W^{1,\infty}([0, T]; V_h)$ is the space of Lipschitz continuous functions from $[0, T]$ to $V_h$.

Given $u_{h0}$ and $u_{h1}$ in $V_h$, we look for a function $u_h$ which will satisfy the following functional requirements:

(3.3)
$$u_h \in W^{1,\infty}([0, T]; V_h),$$

(3.4) $$\ddot{u}_h \in M^1([0,T]; V_h).$$

These conditions imply that $\dot{u}(t)$ has bounded variation in time and the limit $\lim_{t \downarrow 0} \dot{u}_h(t) \equiv \dot{u}_h(0^+)$ exists. The initial conditions are satisfied in the following sense:

(3.5) $$u_h(0) = u_{h0},$$

(3.6)
$$(u_{h1} - \dot{u}_h(0^+), v_h - u_h(0)) \leq 0, \quad \forall v_h \in K_h,$$
$$|\dot{u}_h(0^+)| = |u_{1h}|.$$

Condition (3.6) means that $\dot{u}_h(0^+)$ satisfies a variational inequality *and* an energy condition. In the language of convex analysis, the set of vectors which satisfy the first relation of (3.6) is a translate of the opposite of the orthogonal cone at $u_h(0)$ to $K_h$.

Finally, $u_h$ satisfies an evolution variational inequality given by

(3.7)
$$\left\{ \begin{array}{l} u_h(t) \in K_h \quad \forall t \in [0,T], \\[2mm] \displaystyle\int_0^T \left( (\ddot{u}_h(t), v_h(t) - u_h(t)) + a(u_h(t), v_h(t) - u_h(t)) \right. \\[4mm] \hspace{3cm} \left. - (f, v_h(t) - u_h(t)) \right) dt \geq 0 \\[2mm] \hspace{2cm} \text{for all continuous } v_h \text{ with values in } K_h. \end{array} \right.$$

Relation (3.7) can be written in a slightly different but equivalent form: define an operator $A_h$ from $V_h$ to itself by

$$(A_h u_h, v_h) = a(u_h, v_h), \quad \forall v_h \in V_h,$$

and a maximal monotone operator $\partial \phi_h$ (see [2], [7], [9]) by

(3.8) $$\partial \phi_h(u_h) = \left\{ \begin{array}{ll} \{0\} & \text{if } u_h \in \text{int}(K_h), \\[2mm] \{x \,/\, (x, v_h - u_h) \leq 0, \forall v_h \in K_h\} & \text{if } u_h \in \partial K_h, \\[2mm] \varnothing & \text{otherwise.} \end{array} \right.$$

Then relation (3.7) can be written as

(3.9) $$\ddot{u}_h + A_h u_h + \partial \phi_h(u_h) \ni f_h.$$

Of course, $f_h(t)$ is the projection of $f(t)$ on $V_h$ with respect to the scalar product $(\cdot, \cdot)$.

We know from [10] that problem (3.9) together with conditions (3.1)–(3.6) possesses a solution which moreover satisfies the energy relation

(3.10) $$|\dot{u}_h(t^{\pm})|^2 + a(u_h(t), u_h(t)) = |u_{h1}|^2 + a(u_{h0}, u_{h0}) + 2 \int_0^t (f(s), \dot{u}(s))\, ds.$$

This relation means that when the constraint imposes a jump in velocity, the magnitude of the velocity vector is conserved. Even among the solutions which conserve the energy, there is generally no uniqueness, as was shown in [10].

We denote by $u_h$ an arbitrary solution of (3.3)–(3.7) which satisfies (3.10). We choose initial data $u_{0h}$ and $u_{1h}$ which satisfy

(3.11) $$\lim_{h \downarrow 0}(\|u_{0h} - u_0\| + |u_{h1} - u_1|) = 0.$$

The convergence theorem is the following

THEOREM 3.1. *Assume that (3.11) holds. Then, the sequence $u_h$ converges for the strong topology of $\mathscr{X}_2$ to the unique solution of (2.8)–(2.10).*

*Proof.* The stability of the approximations $u_h$ is an immediate consequence of the energy relation (3.10) and of the subsequent Gronwall inequality which yields

$$\sup(|\dot{u}_h(t)| + \|u_h(t)\|) \leq C\left(\max_h(|u_{1h}| + \|u_{0h}\|), T\right), \quad \forall h \leq h_0.$$

Moreover, it is obvious that $u_h$ belongs to $\mathscr{K}$. We can extract a subsequence, again denoted by $u_h$, such that

$$u_h \rightharpoonup u \quad \text{in } \mathscr{X}_\infty \text{ weak } * .$$

The injection $C^{0,1/2}(Q_T) \hookrightarrow C^{0,\alpha}(Q_T)$ is compact for every $\alpha$ strictly smaller than $1/2$. Therefore, $u$ belongs to $\mathscr{K}$.

In order to prove that the limit $u$ satisfies (2.10), it is necessary to take convenient test functions. The elements of $\mathscr{K}$ are not smooth enough in time, and they have to be approximated before being projected onto $V_h$. This projection does not conserve the constraint at $x = L$, and therefore, the elements of $\mathscr{K}$ need another approximation in order to satisfy the constraint strictly. More precisely, let $v$ be an element of $\mathscr{K}$ which is equal to $u$ for $t \geq T - \varepsilon$. For $\eta \leq \varepsilon/4$, define

$$(3.12) \quad v^\eta(x,t) = \begin{cases} u(x,t) + \frac{1}{\eta} \int_t^{t+\eta} (v - u)(x,s)\, ds + k(\eta)x\phi(t) & \text{if } t \leq T - \eta, \\ u(x,t) & \text{if } t \geq T - \eta. \end{cases}$$

The function $\phi$ is nonnegative and smooth; it is equal to one on $[0, T - \varepsilon/2]$, and it vanishes on $[T - \varepsilon/4, T]$. The parameter $k(\eta)$ is chosen as follows: as $u$ belongs to $\mathscr{X}_\infty$, we have the inequality (see Lemma 2.5)

$$\left| u(L,t) - \frac{1}{\eta} \int_t^{t+\eta} u(L,s)\, ds \right| \leq \frac{1}{\eta} \int_t^{t+\eta} |u(t) - u(s)|\, ds$$

$$\leq C[\![u]\!]_\infty \frac{1}{\eta} \int_0^\eta \sqrt{s}\, ds = \frac{2}{3} C[\![u]\!]_\infty \sqrt{\eta}.$$

Here, $C$ is the norm of the injection $\mathscr{X}_\infty \to L^2(Q_T)$. We have the inequality, for $t \leq T - \varepsilon/2$,

$$v^\eta(L,t) \geq \frac{1}{\eta} \int_t^{t+\eta} v(L,s)\, ds - \frac{2}{3} C[\![u]\!]_\infty \sqrt{\eta} + k(\eta)L\phi(t).$$

If we choose

$$(3.13) \qquad\qquad k(\eta) = \frac{2C}{L}[\![u]\!]_\infty \sqrt{\eta},$$

we will be sure that

$$(3.14) \qquad\qquad v^\eta(L,t) \geq k_0 + C[\![u]\!]_\infty \sqrt{\eta} \quad \text{for } t \leq T - \frac{\varepsilon}{2}.$$

It is not difficult to check that, for $t$ in $[T - \varepsilon/2, T - \eta]$,

$$v^\eta(x,t) = u(x,t) + k(\eta)x\phi(t),$$

so that $v^\eta$ belongs to $\mathscr{K}$. On the other hand, $v^\eta$ belongs to $L^\infty(0,T;V)$ because the time integration has a smoothing effect.

We denote by $Q_h$ the projection onto $V_h$, orthogonal with respect to the scalar product of $H$. The sequence $Q_h$ converges in the strong operator topology of $H$ to the identity, and therefore, thanks to the Sobolev injections, there exists a sequence $\gamma_h$ converging to zero as $h$ tends to zero such that

$$\|Q_h z - z\|_{C^0} \leq \gamma_h \|z\|, \quad \forall z \in V.$$

Moreover, there exists a positive constant $C$ such that

(3.15)                              $$\|Q_h v\| \leq C\|v\|, \quad \forall v \in V.$$

This property is proved by a classical computation.

The test function which will be used in (3.7) is

$$v_h(t) = u_h(t) + Q_h(v^\eta(t) - u(t));$$

by a continuity argument, and the previous construction of $v^\eta$, $v_h(t)$ is an element of $K_h$, for all $t$, and for $h$ small enough. The rest of the proof is standard. $\square$

3.2. *Convergence of Fully Discretized Finite Element Schemes.* With $V_h$ as in (3.1) and $K_h$ as in (3.2), we define a fully discretized scheme

(3.16)
$$\begin{cases} u_h^{k+1} \in K_h; \\ \left(\dfrac{u_h^{k+1} - 2u_h^k + u_h^{k-1}}{\Delta t^2}, v - u_h^{k+1}\right) + a(u_h^k, v - u_h^{k+1}) \geq \left(f_h^k, v - u_h^{k+1}\right), \\ \hspace{8cm} \forall v \in K_h, \end{cases}$$

with $f_h^k$ a suitable discretization of $f$ and initial conditions $u_h^0$ and $u_h^1$ such that

(3.17)                    $$\lim_{h \downarrow 0} \left(\|u_h^0 - u_0\| + \left|\frac{u_h^1 - u_h^0}{\Delta t} - u_1\right|\right) = 0.$$

Using the notation in (3.8), relation (3.16) can be written equivalently as

(3.18)                    $$\frac{u_h^{k+1} - 2u_h^k + u_h^{k-1}}{\Delta t^2} + A_h u_h^k + \partial\phi_h(u_h^{k+1}) \ni f_h^k.$$

The scheme (3.16) is implicit in the constraint. It is equivalent to minimizing a coercive and twice differentiable functional on a convex set. Therefore, it defines a unique $u_h^k$ at each step.

The stability condition will use the constant $\kappa_h$ defined by

(3.19)                    $$\kappa_h = \sup_{u \in V_h \setminus \{0\}} \frac{(A_h u, u)}{|u|^2}.$$

THEOREM 3.2. *Assume that there exists a strictly positive number $\alpha$ such that*

(3.20)                    $$\Delta t \leq 2\sqrt{\frac{1-\alpha}{\kappa_h}}.$$

*Then, under condition* (3.17), *the numerical scheme* (3.16) *converges to the unique solution of* (2.8)–(2.10) *as $h$ and $\Delta t$ converge to zero.*

*Proof.* Let us first prove the stability: if we let $v = u_h^{k-1}$ in (3.16), we can write (3.16) as

(3.21)
$$a(u_h^k, u_h^{k+1}) + \left|\frac{u_h^{k+1} - u_h^k}{\Delta t}\right|^2$$
$$\leq a(u_h^k, u_h^{k-1}) + \left|\frac{u_h^{k-1} - u_h^k}{\Delta t}\right|^2 + (f_h^k, u_h^{k+1} - u_h^{k-1}).$$

Consider the quadratic form over $V_h \times V_h$ defined by

$$R(u,v) = (A_h u, v) + \left| \frac{u-v}{\Delta t} \right|^2,$$

which appears twice in (3.21). We shall show that it is positive definite if condition (3.20) holds. We have the identity

$$(A_h u, v) = \frac{1}{4}(A_h(u+v), u+v) - \frac{1}{4}(A_h(u-v), u-v)$$

and thus

$$R(u,v) \geq \frac{1}{4}(A_h(u+v), u+v) + \left| \frac{u-v}{\Delta t} \right|^2 \left( 1 - \frac{\kappa_h \Delta t^2}{4} \right).$$

Thanks to (3.20),

(3.22) $$R(u,v) \geq \frac{1}{4}(A_h(u+v), u+v) + \alpha \left| \frac{u-v}{\Delta t} \right|^2.$$

If we perform a discrete integration over (3.21), we obtain

$$R(u_h^k, u_h^{k+1}) \leq R(u_h^0, u_h^1) + \sum_{l=1}^{k} \left( f_h^l, \frac{u_h^{l+1} - u_h^{l-1}}{\Delta t} \right) \Delta t.$$

Therefore,

$$\frac{1}{4}\|u_h^{k+1} + u_h^k\| + \alpha \left| \frac{u_h^{k+1} - u_h^k}{\Delta t} \right|^2$$

$$\leq R(u^0, u^1) + \sum_{l=1}^{k} |f^l|^2 \Delta t + 2 \sum_{l=0}^{k} \left| \frac{u^{l+1} - u^l}{\Delta t} \right|^2 \Delta t.$$

This is a discrete Gronwall inequality, which can be integrated readily [5, Lemme 4.1, p. 76], and yields, if $M$ is the largest integer such that $Mh \leq T$,

(3.23) $$\frac{1}{4}\|u_h^{k+1} + u_h^k\| + \alpha \left| \frac{u_h^{k+1} - u_h^k}{\Delta t} \right|^2 \leq C \left( T, R(u_h^0, u_h^1), \sum_{1 \leq l \leq M} |f_h^l|^2 \Delta t \right).$$

Define an interpolation $u_h$ by

$$u_h(x,t) = u_h^k(x) \frac{(k+1)\Delta t - t}{\Delta t} + u_h^{k+1}(x) \frac{t - k\Delta t}{\Delta t} \quad \text{for } k\Delta t \leq t \leq (k+1)\Delta t.$$

Relation (3.23) implies that we can extract from the sequence $(u_h)_h$ a subsequence (again denoted by $(u_h)_h$) such that

$$u_h \rightharpoonup u \quad \text{in } L^\infty(0,T;V) \text{ weak } *,$$

$$\frac{du_h}{dt} \rightharpoonup u \quad \text{in } L^\infty(0,T;H) \text{ weak } *,$$

$$u_h \to u \quad \text{in } C^{0,\beta}(Q_T), \quad \forall \beta < \frac{1}{2}.$$

Now, we choose $v^\eta$ as in (3.12), and we let

(3.24) $$v_h^k = u_h^{k+1} + Q_h\big(v^\eta(k\Delta t) - u(k\Delta t)\big).$$

If we substitute this value for $v$ in (3.16) and perform a discrete time integration, we obtain

$$
\begin{aligned}
&- \left( \frac{u_h^1 - u_h^0}{\Delta t}, v_h^0 - u_h^1 \right) - \sum_{k=1}^{M-1} \left( \frac{u_h^k - u_h^{k-1}}{\Delta t}, \frac{v_h^k - u_h^{k+1} - v_h^{k-1} + u_h^k}{\Delta t} \right) \Delta t \\
&+ \sum_{k=1}^{M-1} a(u_h^k, v_h^k - u_h^{k+1}) \Delta t \geq \sum_{k=1}^{M-1} \left( f_h^k, v_h^k - u_h^{k+1} \right) \Delta t.
\end{aligned}
$$
(3.25)

The passage to the limit in this expression is obvious. In order to have strong convergence in $\mathscr{H}_2$, it is enough to show that the total energy of $u_h$ converges to the total energy of $u$. This is done by a discrete integration of (3.21). $\quad\square$

Let us now compute the value of $\kappa_h$ for uniform $P_1$ finite elements, corresponding to a step $\Delta x = L/n$. It is a classical result that

$$
\|u\|^2 \leq \frac{12|u|^2}{\Delta x^2},
$$

so that $\kappa_h = 12/\Delta x^2$. Therefore, the stability condition (3.20) becomes

$$
\Delta t \leq \frac{\Delta x \sqrt{1-\alpha}}{\sqrt{3}}.
$$

## 4. Explicit Difference Schemes.

4.1. *Notations and Description of the Family of Schemes.* Let $n$ be an integer, let $h = \Delta x = L/n$ be the space step, and let $\Delta t$ be the time step. We denote by $U_j^k$ the solution of the finite difference scheme defined as follows:

$$
(4.1) \quad
\begin{cases}
U_j^{k+1} = 2U_j^k - U_j^{k-1} + \dfrac{\Delta t^2}{\Delta x^2} \left( U_{j+1}^k - 2U_j^k + U_{j-1}^k \right) + \Delta t^2 F_j^k \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for } 1 \leq j \leq n-1, \\
U_n^{k+1} = \max\left[ k_0, 2U_n^k - U_n^{k-1} + \dfrac{\Delta t^2}{\Delta x^2} \left( U_{n-1}^k - U_n^k \right) + \Delta t^2 F_n^k \right],
\end{cases}
$$

where $F_j^k$ is a suitable discretization of $f$. By convention, $U_0^k = 0$.

If we define an interpolation in the space of $U_j^k$, and a scalar product which is an approximation of the $L^2$ scalar product, this family of schemes admits a variational formulation. Namely, let

$$
g(x) = \max(1 - |x|,\ 0),
$$
(4.2)

and

$$
u_h^k(x) = \sum_{j=1}^n U_j^k\, g\left( \frac{x - j\Delta x}{\Delta x} \right).
$$
(4.3)

Then $u_h^k$ is an element of $V_h$, the space of uniform $P_1$ finite elements with nodes at the points $jh$, for $0 \leq j \leq n$. The function $f_h^k$ is defined from $F_h^k$ in a similar fashion. Let $u$ and $v$ be of the form

$$
u = \sum_{j=1}^n U_j\, g\left( \frac{x - j\Delta x}{\Delta x} \right), \qquad v = \sum_{j=1}^n V_j\, g\left( \frac{x - j\Delta x}{\Delta x} \right);
$$

we define an alternative scalar product over $V_h$ by

$$(4.4) \qquad s_h(u, v) = \sum_{j=1}^{n} U_j V_j \Delta x.$$

The properties of $s_h$ are summarized in the following result:

LEMMA 4.1. *For all $u$ and $v$ in $V_h$ of the form*

$$u = \sum_{j=1}^{n} g\left(\frac{x - j\Delta x}{r}\right) U_j, \qquad v = \sum_{j=1}^{n} g\left(\frac{x - j\Delta x}{\Delta x}\right) V_j,$$

*the scalar product $s_h$ satisfies the following identity:*

$$(4.5) \qquad s_h(u, v) = (u, v) + \frac{\Delta x^2}{6} a(u, v) + \Delta x \frac{U_n V_n}{2}.$$

*Moreover, we have the inequality*

$$(4.6) \qquad a(u, u) \le 4 \frac{s_h(u, u)}{\Delta x^2}, \quad \forall u \in V_h.$$

*Proof.* We observe that

$$s_h(u, u) - (u, u) = \sum_{j=0}^{n-1} \frac{1}{2}(U_j^2 + U_{j+1}^2)\Delta x + \frac{1}{2}U_n^2 \Delta x$$

$$- \sum_{j=0}^{n-1} \frac{1}{3}(U_j^2 + U_j U_{j+1} + U_{j+1}^2)\Delta x$$

$$= \frac{1}{6}\sum_{j=0}^{n-1}(U_{j+1} - U_j)^2 \Delta x + \frac{1}{2}U_n^2 \Delta x.$$

This shows that

$$s_h(u, u) = (u, u) + \frac{\Delta x^2}{6}\|u\|^2 + \Delta x \frac{U_n^2}{2}.$$

Relation (4.5) follows immediately by differentiation. The other relation is a consequence of the elementary inequality

$$|U_{j+1} - U_j|^2 \le 2(U_{j+1}^2 + U_j^2).$$

Adding these relations with respect to $j$, we obtain

$$\frac{1}{\Delta x}\sum_{j=1}^{n-1}|U_{j+1} - U_j|^2 \le \frac{2}{\Delta x^2}\sum_{j=1}^{n-1}(U_{j+1}^2 + U_j^2)\Delta x \le \frac{4}{\Delta x^2}s_h(u, u).$$

This proves the result. $\square$

LEMMA 4.2. *The finite difference scheme (4.1) is equivalent to the variational inequality*

$$u_h^{k+1} \in K_h,$$

$$(4.7) \qquad s_h\left(\frac{u_h^{k+1} - 2u_h^k + u_h^{k-1}}{\Delta t^2}, v - u_h^{k+1}\right) + a\left(u_h^k, v - u_h^{k+1}\right)$$

$$\ge s_h\left(f_h^k, v - u_h^{k+1}\right), \quad \forall v \in K_h.$$

*Proof.* According to (3.2), the convex set $K_h$ is the set

$$K_h = \{u \in V_h \, / \, U_n = u(L) \geq k_0\}.$$

If we substitute $v$ in (4.7) by $u + \phi$, with $\phi(L) = 0$, then

$$s_h\left(\frac{u_h^{k+1} - 2u_h^k + u_h^{k-1}}{\Delta t^2}, \phi\right) + a\left(u_h^k, \phi\right) \geq s_h\left(f_h^k, \phi\right),$$

which is clearly equivalent to the first relation in (4.1). Take now $v$ such that

$$V_j = U_j, \quad \forall j \leq n-1 \text{ and } V_n \geq k_0 \, ;$$

then we find, using the explicit form of $a(u, v)$ in terms of the $U_j$ and $V_j$,

$$\frac{U_n^{k+1} - 2U_n^k + U_n^{k-1}}{\Delta t^2}\left(V_n - U_n^{k+1}\right) + \frac{U_n^k - U_n^{k-1}}{\Delta x^2}\left(V_n - U_n^{k+1}\right)$$
$$\geq F_n^k\left(V_n - U_n^{k+1}\right), \quad \forall V_n \geq k_0.$$

It is straightforward to check that this relation is equivalent to the second relation of (4.1). $\square$

To make things more precise, we shall settle for the following approximations of the exterior forces and of the initial data:

$$(4.8) \qquad\qquad U_j^0 = u_0(j\Delta x),$$

$$(4.9) \qquad \begin{cases} U_j^1 = U_j^0 + \dfrac{\Delta t}{\Delta x} \displaystyle\int_{(j-1/2)\Delta x}^{(j+1/2)\Delta x} u_1(x)dx, & \text{if } 1 \leq j \leq n-1, \\[3mm] U_n^1 = U_n^0 + \dfrac{2\Delta t}{\Delta x} \displaystyle\int_{(n-1/2)\Delta x}^{n\Delta x} u_1(x)dx, \end{cases}$$

$$(4.10) \qquad\qquad F_j^k = \frac{1}{\Delta t \Delta x} \int_{(j-1)\Delta x}^{j\Delta x} \int_{k\Delta t}^{(k+1)\Delta t} f(x, t)\,dx dt.$$

In the remainder of this section, we shall consider that $\lambda = \Delta t / \Delta x$ is a fixed number belonging to $[0, 1]$; we shall first prove the convergence of the scheme in the noncharacteristic case ($\lambda < 1$), and then in the characteristic case ($\lambda = 1$).

4.2. *Convergence in the Noncharacteristic Case.*

PROPOSITION 4.3. *Let $\lambda$ be strictly less than 1. Assume that the initial data and the right-hand side are discretized according to (4.8)–(4.10). Then there exists a constant $C$ which depends only on $|u_0|$, $\|u_1\|$, $|f|_{L^2(Q_T)}$ and $T$ such that*

$$\|u_h^k\| + s_h\left(\frac{u_h^{k+1} - u_h^k}{\Delta t}, \frac{u_h^{k+1} - u_h^k}{\Delta t}\right)^{1/2} \leq C,$$

*for all $k$ less than or equal to $T/\Delta t$.*

*Proof.* With exactly the same technique as in the proof of stability in Theorem 3.2, we consider the expression

$$F'(u, v) = a(u, v) + s_h\left(\frac{v-u}{\Delta t}, \frac{v-u}{\Delta t}\right).$$

From inequality (4.6) we deduce that

$$R'(u,v) = \frac{1}{4}a(u+v, u+v) - \frac{1}{4}a(u-v, u-v) + s_h\Big(\frac{v-u}{\Delta t}, \frac{v-u}{\Delta t}\Big)$$

$$\geq \frac{1}{4}a(u+v, u+v) + s_h\Big(\frac{u-v}{\Delta t}, \frac{u-v}{\Delta t}\Big)(1-\lambda^2).$$

Therefore, under the assumptions of the proposition, $R'$ is a positive definite quadratic form over $V_h \times V_h$ which satisfies the coercivity inequality

$$R'(u,v) \geq C(\lambda)\big(a(u,u) + a(v,v)\big),$$

where $C(\lambda)$ is some positive constant. Then we perform a discrete time integration; we use a discrete Gronwall inequality, and we obtain the result, exactly as in the proof of stability in Theorem 3.2.   □

We interpolate $u_h^k$ by letting

$$u_h(x,t) = u_h^k(x)\frac{(k+1)\Delta t - y}{\Delta t} + u_h^{k+1}(x)\frac{t - \Delta t}{\Delta t}.$$

We can now state a convergence result.

THEOREM 4.4. *Assume that the initial data and the right-hand side are discretized according to* (4.8)–(4.10), *and that* $\lambda$ *is strictly less than* 1. *Then, the sequence* $u_h$ *converges to the unique solution* $u$ *of* (2.8)–(2.10) *for the strong topology of* $\mathscr{X}_2$.

*Proof.* From Proposition 4.2, we can see that we can extract a converging subsequence, again denoted by $u_h$, such that

$$u_h \rightharpoonup u \quad \text{in } L^\infty(0,T;V) \text{ weak } *,$$

$$\frac{du_h}{dt} \rightharpoonup u \quad \text{in } L^\infty(0,T;H) \text{ weak } *,$$

$$u_h \to u \quad \text{in } C^{0,\beta}(Q_T), \quad \forall \beta < \frac{1}{2}.$$

Clearly, $u$ belongs to $K$. We define $v^\eta$ as in (3.12), and $v_h^k$ as in (3.24) by

$$v_h^k = u_h^{k+1} + Q_h(v^\eta(k\Delta t) - u(k\Delta t)).$$

We substitute $v$ by $v_h^k$ in (4.5); a discrete integration in time yields

(4.11)
$$-s_h\Big(\frac{u_h^1 - u_h^0}{\Delta t}, v_h^0 - u_h^1\Big)$$
$$-\sum_{k=1}^{M-1} s_h\Big(\frac{u_h^k - u_h^{k-1}}{\Delta t}, \frac{v_h^k - u_h^{k+1} - v_h^{k-1} + u_h^k}{\Delta t}\Big)\Delta t$$
$$+ \sum_{k=1}^{M-1} a\big(u_h^k, v_h^k - u_h^{k+1}\big)\Delta t \geq \sum_{k=1}^{M-1} s_h\big(f_h^k, v_h^k - u_h^{k+1}\big)\Delta t.$$

The only difference between (4.11) and (3.25) is that the scalar product $(\cdot, \cdot)$ is replaced by the scalar product $s_h(\cdot, \cdot)$. Therefore, if we substitute in (4.11) the

scalar product $s_h$ by the regular scalar product, we commit an error given by

$$\frac{\Delta x^2}{6} a\Big(\frac{u_h^1 - u_h^0}{\Delta t}, v^0 - u^1\Big) + \frac{\Delta x}{2}\Big(\frac{U_n^1 - U_n^0}{\Delta t}\Big)\big(V_n^0 - U_n^1\big)$$

$$+ \frac{\Delta x^2}{6} \sum_{k=1}^{M-1} a\Big(\frac{u_h^k - u_h^{k-1}}{\Delta t}, \frac{v_h^k - u_h^{k+1} - v_h^{k-1} + u_h^k}{\Delta t}\Big)\Delta t$$

(4.12)
$$+ \frac{\Delta x}{2} \sum_{k=1}^{M-1} \Big(\frac{U_n^k - U_n^{k-1}}{\Delta t}\Big)\Big(\frac{V_n^k - U_n^{k+1} - V_n^{k-1} + U_n^k}{\Delta t}\Big)\Delta t$$

$$+ \frac{\Delta x^2}{6} \sum_{k=1}^{M-1} a\big(f_h^k, v_h^k - u_h^{k+1}\big)\Delta t + \frac{\Delta x}{2} \sum_{k=1}^{M-1} F_n^k\big(V_n^k - U_n^{k+1}\big)\Delta t$$

$$= T_1 + T_2 + T_3 + T_4 + T_5 + T_6.$$

We have to estimate all the terms $T_i$, $i = 1, \dots, 6$. For the term $T_1$, we extend $u_1$ to $[L, 2L]$ by letting

$$u_1(L + x) = u_1(L - x)$$

and we observe that (4.9) can be written for $1 \le j \le n$ as

$$\frac{U_j^1 - U_j^0}{\Delta t} = \frac{1}{\Delta x} \int_{(j-1/2)\Delta x}^{(j+1/2)\Delta x} u_1(s)\, ds,$$

so that

$$\left|\frac{U_j^1 - U_j^0}{\Delta t}\right| \le \frac{1}{\sqrt{\Delta x}} \left[\int_{(j-1/2)\Delta x}^{(j+1/2)\Delta x} \big|u_1(s)\big|^2\, ds\right]^{1/2}.$$

This implies that

$$\sum_{j=1}^{n} \left|\frac{U_j^1 - U_j^0}{\Delta t}\right|^2 \Delta x \le \sum_{j=1}^{n} \frac{1}{\Delta x}\Delta x \int_{(j-1/2)\Delta x}^{(j+1/2)\Delta x} \big|u_1(s)\big|^2\, ds \le 2\big|u_1\big|^2.$$

On the other hand, from relation (4.6), we get

$$\left\|\frac{u_h^1 - u_h^0}{\Delta t}\right\| \le \frac{2\sqrt{2}|u_1|}{\Delta x}.$$

Therefore,

(4.13)
$$\big|T_1\big| \le \frac{\sqrt{2}\Delta x}{3}\, |u_1|\, \|v^0 - u^1\|.$$

The next term is estimated by

$$\big|T_2\big| \le C\sqrt{\frac{\Delta x}{2}}\, |u_1|\, |V_n^0 - U_n^1|.$$

It is clear that $U_n^1$ and $V_n^0$ are bounded independently of $h$. Thus we have

(4.14)
$$\big|T_2\big| = O(\sqrt{\Delta x}).$$

For the third term, we observe that

$$\|v_h^k - u_h^{k+1} - v_h^{k-1} + u_h^k\|$$
$$= \|Q_h\big(v^\eta(k\Delta t) - u(k\Delta t) - v^\eta((k-1)\Delta t) - u((k+1)\Delta t)\big)\|$$
$$\le \|v^\eta(k\Delta t) - u(k\Delta t) - v^\eta((k-1)\Delta t) - u((k+1)\Delta t)\|.$$

But

$$v^\eta(k\Delta t) - u(k\Delta t) = \frac{1}{\eta} \int_{k\Delta t}^{k\Delta t + \eta} (v - u)(x, s)\, ds + k(\eta) x \phi(k\Delta t).$$

Therefore, for $\Delta t$ small enough,

$$\|v^\eta(k\Delta t) - u(k\Delta t)\| \le \frac{1}{\eta} \sqrt{\Delta t}\, [\![v - u]\!]_2 + Lk(\eta)|\phi'|_\infty \Delta t$$

$$\le \frac{2}{\eta} \sqrt{\Delta t}\, [\![v - u]\!]_2.$$

The estimate for $T_3$ is

(4.15)
$$|T_3| \le \frac{(M-1)\Delta x^2}{6} \frac{4}{\eta\sqrt{\Delta t}} [\![v - u]\!]_2 \max_k |U_h^k|$$

$$\le \frac{2\sqrt{\Delta t}}{3\lambda^2 \eta} T [\![v - u]\!]_2 \max_k \|u_h^k\| = O(\sqrt{\Delta x}).$$

The estimate for $T_4$ relies on the inequality

$$|V_n^k - U_n^{k+1}| \le C \frac{\sqrt{\Delta t}}{\eta} \left[ \int_{k\Delta t}^{k\Delta t + \eta} \|(v - u)(x, s)\|^2\, ds \right]^{1/2},$$

from which we deduce

$$|T_4| \le \Delta x\, C \frac{\sqrt{\Delta t}}{\eta} \max_k \|u_k\| \sum_{k=1}^{M-1} \left\{ \left[ \int_{k\Delta t}^{k\Delta t + \eta} \|(v - u)(x, s)\|^2\, ds \right]^{1/2} \right.$$

$$\left. + \left[ \int_{(k-1)\Delta t}^{(k-1)\Delta t + \eta} \|(v - u)(x, s)\|^2\, ds \right]^{1/2} \right\}$$

$$\le \Delta x\, C \frac{\sqrt{\Delta t}}{\eta} \max_k \|u_k\| \sqrt{M-1} \left\{ \left[ \sum_{k=1}^{M-1} \int_{k\Delta t}^{k\Delta t + \eta} \|(v - u)(x, s)\|^2\, ds \right]^{1/2} \right.$$

$$\left. + \left[ \sum_{k=1}^{M-1} \int_{(k-1)\Delta t}^{(k-1)\Delta t + \eta} \|(v - u)(x, s)\|^2\, ds \right]^{1/2} \right\}.$$

If $m$ is an integer such that $m\Delta t \le \eta < (m+1)\Delta t$, we have finally the inequality

(4.16)
$$|T_4| \le 2C\Delta x \frac{\sqrt{\Delta t}}{\eta} \max_k \|u_k\| \sqrt{\frac{T}{\Delta t}} \sqrt{\frac{\eta}{\Delta t} + 1}\, [\![v - u]\!]_2 = O(\sqrt{\Delta x}).$$

For the last two terms, we have the easy estimates

(4.17)
$$|T_5| = O(\Delta x),$$

(4.18)
$$|T_6| = O(\sqrt{\Delta x}).$$

Finally, adding estimates (4.13) to (4.18), we can see that the error (4.12) is $O(\sqrt{\Delta x})$, and the passage to the limit is performed in the same fashion as in the proof of Theorem 3.2.    □

4.3. *Convergence in the Characteristic Case.* We have first an energy estimate which is analogous to the noncharacteristic case:

PROPOSITION 4.5. *Let $\lambda = 1$ and*

$$R'(u,v) = a(u,v) + s_h\left(\frac{u-v}{\Delta x}, \frac{u-v}{\Delta x}\right).$$

*Let $U_j^k$ be defined by* (4.1), *with data* (4.8)–(4.10). *The interpolation $u_h^k$ of $U_j^k$ is given by* (4.3). *Then $R'(u_h^{k+1}, u_h^k)$ is bounded for $kh \leq T$ by a constant depending only on the data $u_0$, $u_1$, and $f$ and on $T$.*

*Proof.* If $\Delta t = \Delta x$, the quadratic form $R'$ defined above is no longer coercive on $V_h \times V_h$, uniformly in $h$; nevertheless, it is still positive definite. Its explicit expression is indeed equal to

$$R'(u,v) = \sum_{j=0}^{n-1} \frac{(U_{j+1} - U_j)(V_{j+1} - V_j)}{\Delta x} + \sum_{j=1}^{n} \frac{(V_j - U_j)^2}{\Delta x}.$$

Elementary manipulations give a form which is easier to handle, namely

$$(4.19) \qquad R'(u,v) = \sum_{j=0}^{n-1} \frac{(U_{j+1} - V_j)^2 + (U_j - V_{j+1})^2}{2\Delta x} + \frac{(U_n - V_n)^2}{2\Delta x}.$$

In (4.19) it is easy to see that $R'$ is positive definite. Moreover, (4.7) becomes, when $v$ is substituted by $u_h^{k-1}$,

$$R'(u_h^k, u_h^{k+1}) \leq R'(u_h^{k-1}, u_h^k) + s_h\left(f_h^k, \frac{u_h^{k+1} - u_h^{k-1}}{\Delta x}\right)\Delta x.$$

But we have the inequalities

$$s_h\left(f_h^k, \frac{u_h^{k+1} - u_h^{k-1}}{\Delta x}\right) \leq \frac{1}{2}s_h(f_h^k, f_h^k) + \frac{1}{2}s_h\left(\frac{u_h^{k+1} - u_h^{k-1}}{\Delta x}, \frac{u_h^{k+1} - u_h^{k-1}}{\Delta x}\right),$$

and

$$\sum_{j=1}^{n} \Delta x \left|\frac{U_j^{k+1} - U_j^{k-1}}{\Delta x}\right|^2$$

$$\leq 2\sum_{j=1}^{n} \Delta x \left|\frac{U_j^{k+1} - U_{j-1}^k}{\Delta x}\right|^2 + 2\sum_{j=1}^{n} \Delta x \left|\frac{U_j^{k-1} - U_{j-1}^k}{\Delta x}\right|^2.$$

Therefore,

$$s_h\left(f_h^k, \frac{u_h^{k+1} - u_h^{k-1}}{\Delta x}\right) \leq R'(u_h^{k+1}, u_h^k) + R'(u_h^k, u_h^{k-1}) + \frac{1}{2}s_h(f_h^k, f_h^k).$$

From a discrete Gronwall estimate we can deduce that

$$(4.20) \qquad\qquad R'(u_h^k, u_h^{k+1}) \leq C(f, T, u^0, u^1).$$

This proves the stated energy estimate. $\square$

The convergence theorem can be formulated as follows:

THEOREM 4.6. *Assume that the initial data and the right-hand side of* (4.1) *are discretized according to* (4.8)–(4.10), *and that $\lambda = 1$. Then the sequence $u_h$ converges to the unique solution $u$ of* (2.8)–(2.10) *in the strong topology of $\mathscr{X}_2$.*

*Proof.* The difficulty in the proof of convergence comes from the fact that the approximate solution does not satisfy simple estimates as in the proof of Theorem 4.4. Because of the presence of staggered grids, we estimate

$$\frac{|U_1^k|^2}{\Delta x} + \sum_{j=0}^{n-2} \frac{\left|U_{j+2}^k - U_j^k\right|^2}{\Delta x} \quad \text{and} \quad \sum_{j=0}^{n} \frac{\left|U_j^{k+2} - U_j^k\right|^2}{\Delta x}.$$

Let $m$ be an integer between 0 and 3, and let $m = 2b_1 + b_0$ be its binary expansion. We define four subgrids $G_h^{(m)}$ by

$$G^{(m)} = \left( [(2\mathbf{N} + b_0) \cap \{0, \dots, n\}] \cup \{0\} \cup \{n\} \right) \times \left( (2\mathbf{N} + b_1) \cup \{0\} \right).$$

It is immediate that for any $m = 0, 1, 2, 3$, and any pair $(j, k)$ in $G_h^{(m)}$ such that $2 \le j \le n - 2$ and $2 \le k$ the following holds:

$$U_j^{k+2} = U_{j-2}^k + U_{j+2}^k - U_j^{k-2}.$$

The grid $G^{(m)}$ defines elementary rectangles with vertices at neighboring points of the grid. We define an interpolation $u_h^{(m)}$ with respect to each grid $G_h^{(m)}$ by requiring $u_h^{(m)}$ to be continuous, to be equal to the discrete solution at the nodes:

$$u_h^{(m)}(j\Delta x, k\Delta x) = U_j^k, \quad \forall (j, k) \in G_h^{(m)},$$

and to coincide with a polynomial of the form $a + bx + ct + dxt$ in each of the elementary rectangles of the grid. Then, from (4.20), $u_h^{(m)}$ is bounded in $\mathscr{X}_\infty$ and by extraction, there is a subsequence, again denoted by $u_h^{(m)}$, such that

$$u_h^{(m)} \rightharpoonup u^{(m)} \quad \text{in } L^\infty(0, T; V),$$

$$\frac{du_h^{(m)}}{dt} \rightharpoonup \frac{du^{(m)}}{dt} \quad \text{in } L^\infty(0, T; H),$$

$$u_h^{(m)} \to u \quad \text{in } C^{0,\beta}(Q_T), \forall \beta < \frac{1}{2}.$$

It is very easy to see that, in the limit,

$$\Box u^{(m)} = f \quad \text{in } Q_T.$$

Moreover, the initial conditions are the same for each of the $u^{(m)}$, and they all satisfy the boundary condition $u^{(m)}(0, t) = 0$. It remains to see that they coincide for $x = L$. First, because of the relation

$$U_j^{k+1} = U_{j-1}^k + U_{j+1}^k - U_j^{k-1},$$

a passage to the limit shows that $u^{(0)} = u^{(3)}$ and $u^{(1)} = u^{(2)}$. Moreover, the relation

$$U_n^{k+1} = \max(k_0, U_n^k + U_{n-1}^k - U_{n-1}^{k-1})$$

yields, by a passage to the limit, if $n + k$ is even,

$$u^{(1)}(L, t) = \max\left(k_0, u^{(0)}(L, t)\right),$$

and if $n + k$ is odd,

$$u^{(0)}(L, t) = \max\left(k_0, u^{(1)}(L, t)\right).$$

Thus, all the $u^{(m)}$ coincide on the boundary $x = L$ of $Q_T$. By classical results on the uniqueness of the solutions of the wave equation, they coincide on all of $Q_T$.

Let us now find a variational inequality satisfied by the $u_h^{(m)}$. We go back to (4.7), and we choose a $v = v^k$ in $K$ such that

$$v^k - u_h^{k+1} = v_*^k - u_h^{k+2} = v_\flat^k - u_h^k = z.$$

From now on, we assume that $n$ is even, and the technical modifications which must be made in the case $n$ odd are left to the reader. We add twice (4.7) at discrete time $k$ to (4.7) at discrete time $k - 1$ and at discrete time $k + 1$. We obtain

$$
(4.21) \qquad
\begin{aligned}
s_h &\left( \frac{u_h^{k+2} - 2u_h^k + u_h^{k-2}}{\Delta t^2}, z \right) + a(u_h^{k-1} + 2u_h^k + u_h^{k+1}, z) \\
&\geq s_h \left( f_h^{k-1} + 2f_h^k + f_h^{k+1}, z \right).
\end{aligned}
$$

Let $\bar{V}_h$ be the subspace of all functions of $V_h$ which are affine on

$$[2j\Delta x, (2j + 2)\Delta x];$$

an element $z$ of $\bar{V}_h$ is of the form

$$z(x) = \sum_{j=0}^{\bar{n}} \bar{Z}_j g\left( \frac{x - 2j\Delta x}{2\Delta x} \right),$$

with $\bar{n} = n/2$. An easy computation shows that, for any $w$ in $V_h$, we have

$$s_h(w, z) = 2\Delta x \sum_{j=1}^{\bar{n}-1} \bar{Z}_j \left( \frac{W_{j-1} + 2W_j + W_{j+1}}{4} \right) + 2\Delta x\, \bar{Z}_{\bar{n}} \left( \frac{2W_n + W_{n-1}}{4} \right),$$

and

$$a(w, z) = \sum_{j=0}^{\bar{n}-1} \frac{(W_{2j+2} - W_{2j})(\bar{Z}_{j+1} - \bar{Z}_j)}{2\Delta x}.$$

Let $\tilde{V}_h$ be the subspace of all the functions of $V_h$ which are affine on $[0, \Delta x]$, on $[(n-1)\Delta x, n\Delta x]$, and all the intervals $[(2j - 1)\Delta x, (2j + 1)\Delta x]$, for $1 \leq j \leq \bar{n} - 1$. We define a mapping from $\tilde{V}_h$ to $\bar{V}_h$ by

$$(Nw)(2j\Delta x) = \begin{cases} [w((2j - 1)\Delta x) + w((2j + 1)\Delta x)]/2 & \text{if } j < n/2 , \\ w((n - 1)\Delta x)/2 & \text{otherwise.} \end{cases}$$

We define a scalar product on $\bar{V}_h$ by

$$\bar{s}_h(u, v) = \sum_{j=1}^{\bar{n}} 2\Delta x\, u(2j\Delta x)\, v(2j\Delta x).$$

With these notations, we can write (4.21) as follows:

$$
\begin{aligned}
\bar{s}_h &\left( \frac{u_h^{(0),k+2} - 2u_h^{(0),k} + u_h^{(0),k-2}}{2\Delta t^2}, z \right) + \bar{s}_h \left( N \frac{u_h^{(2),k+2} - 2u_h^{(2),k} + u_h^{(2),k-2}}{2\Delta t^2}, z \right) \\
&+ a(2u_h^{(0),k} + u_h^{(1),k-1} + u_h^{(1),k+1}, z) \geq s_h(f_h^{k+1} + 2f_h^k + f_h^{k-1}, z).
\end{aligned}
$$

Let $\bar{Q}_h$ be the orthogonal projection on $\bar{V}_h$ with respect to the scalar product $(\cdot, \cdot)$. We define $v_h^k$ by

$$v_h^k = u_h^{k+1} + \bar{Q}_h(v^\eta(k\Delta t) - u(k\Delta t)).$$

Using the same methods as in the noncharacteristic case, and the fact that all the $u^{(m)}$ coincide, we can prove that the limit $u$ satisfies the variational inequality (2.10). By an energy argument, the convergence of the $u_h^{(m)}$ is strong, and therefore, the usual interpolation $u_h$ converges to $u$ strongly in $\mathscr{X}_2$.    □

**5. Numerical Experiments.** We have taken $L = 1$, a contact at $x = 1$ and initial data such that the solution of the linear problem would be a wave propagating towards the positive $x$ direction, and then being reflected. We use the finite difference schemes, either characteristic or subcharacteristic, with initial data which are differentiable or not differentiable.

In order to have a wave travelling to the right, the initial data must satisfy

$$u_1 = -\frac{du_0}{dx}.$$

In the smooth case, the initial position $u_0$ is defined by

$$u_0(x) = \begin{cases} - \left[(x - x_0)^2 - \alpha^2\right]^2 \alpha^{-4} & \text{on } [x_0 - \alpha, x_0 + \alpha], \\ 0 & \text{elsewhere.} \end{cases}$$

In our experiments,

$$\alpha = 0.2, \quad x_0 = 0.6.$$

In order to have a good approximation of a wave travelling to the right, the discretization of the initial data will be

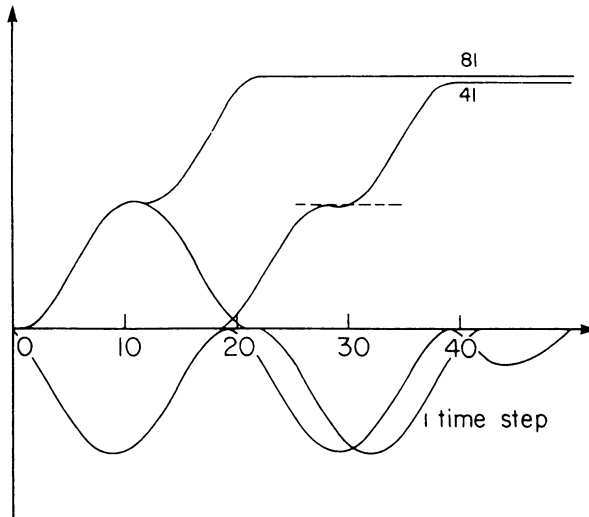$$U_j^0 = u_0(j\Delta x), \qquad U_j^1 = u_0(j\Delta x - \Delta t).$$



FIGURE 2

The first experiment is with a characteristic scheme, and a time step of $1/50$. The results look quite satisfactory, though a careful inspection of the returning wave, after reflection on the side with a unilateral constraint, reveals a small hollow which does not exist in the exact solution (see Figure 2). This small hollow is
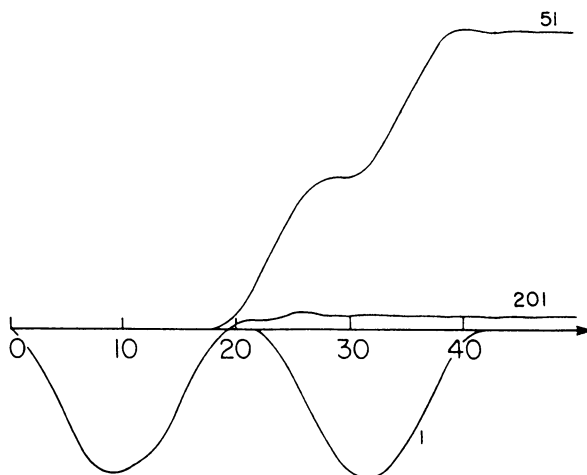
FIGURE 3

due to the effect of the unilateral constraint which implies a change in boundary condition, together with the change of the sense of variation of the wave.

In the second experiment (see Figure 3), with the same space step, and a CFL number of 0.8, the small hollow is not apparent, but the dispersive effect becomes large at later times. If the spatial step is refined to 1/200, without changing the CFL number, the results are substantially better, as in the linear case (see Figure 4).
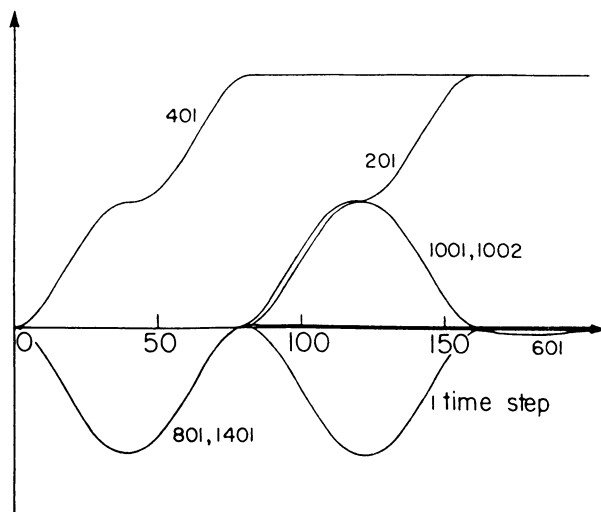


FIGURE 4

The piecewise affine data are

$$u_0(x) = -\max((1 - 5|x - 0.6|), 0).$$

With a space step of 1/50, and a characteristic scheme, we have a notch in the middle of the returning wave; it is too large to make the simulation acceptable;
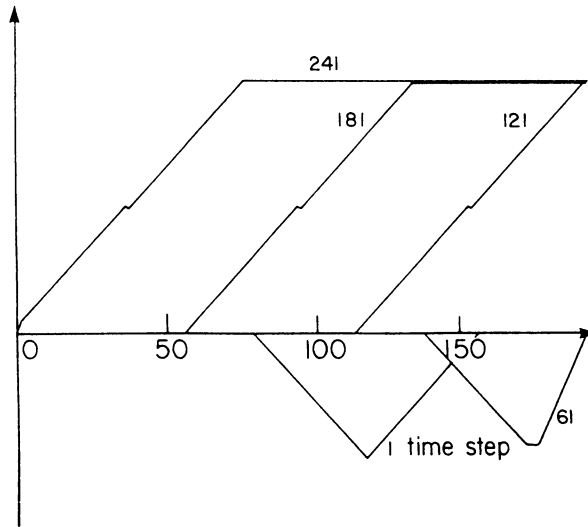
FIGURE 5

with a refined space step of 1/200 (see Figure 5), the notch is still there, but much smaller. Experimentally, the notch is first-order.

These same initial data, and a subcharacteristic scheme with a CFL number of 0.8, lead to noticeable oscillations where the notch was in the characteristic scheme. This is not a surprising result in view of the linear analysis of these schemes. But the dispersive effect of this scheme improves the aspect of the solution with time; see Figure 6.
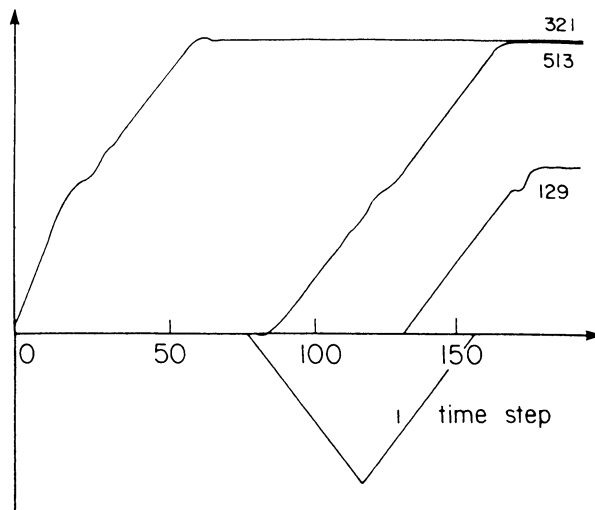


FIGURE 6

The next set of computations (Figure 7) presents the evolution in time of the solution with initial data

$$u_0(x) = \begin{cases} 0 & \text{if } 0 \le x \le 0.5, \\ 50(x - 0.50) & \text{if } 0.50 \le x \le 0.51, \\ 0.5 & \text{if } 0.51 \le x \le 0.60, \\ 50(0.61 - x) & \text{if } 0.60 \le x \le 0.62, \\ -0.5 & \text{if } 0.62 \le x \le 0.72, \\ 50(x - 0.73) & \text{if } 0.72 \le x \le 0.73, \\ 0 & \text{if } 0.73 \le x \le 1. \end{cases}$$
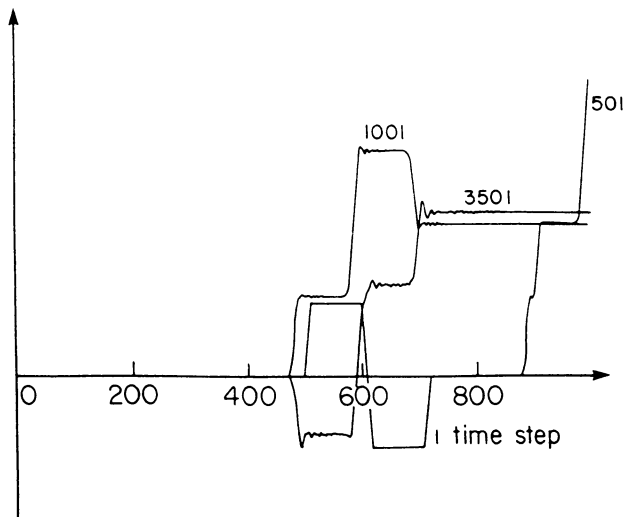


FIGURE 7

This initial condition has quite steep spatial derivatives; with a CFL of 0.08, and a space step of 1/1000, the result has the normally expected oscillations, but is acceptable.

Finally in the last set of computations (Figure 8) we display the evolution of the solution with data

$$u_0(x) = \begin{cases} 0 & \text{if } 0 \le x \le 0.25, \\ \cos(20\pi(x - 0.25)) & \text{if } 0.25 \le x \le 0.75, \\ 0 & \text{if } 0.75 \le x \le 1. \end{cases}$$

We take a space step of 1/200 and a characteristic scheme. The humps are returned one by one by the unilateral constraint.

The choice of a scheme can be governed by the following principles: if the initial data are smooth, the characteristic scheme gives very good results; if the initial data are not smooth, the nonlinear boundary conditions make the computation harder, and create a first-order notch. However, the subcharacteristic scheme gives worse results, even if they look smoother.
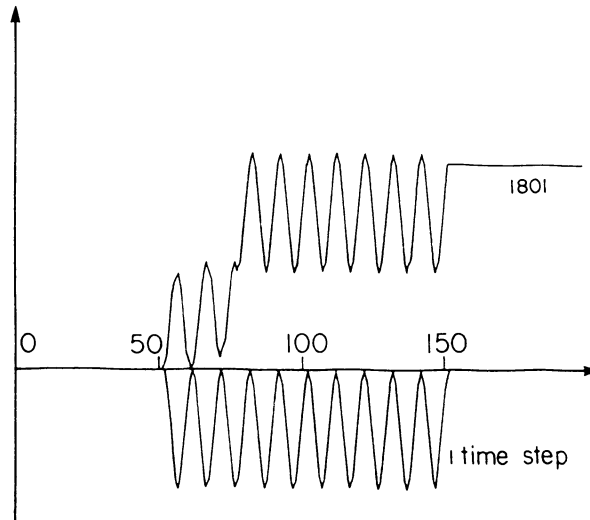
FIGURE 8

Département de Mathématiques
Université Claude-Bernard
69622 Villeurbanne Cedex, France

Department of Computer Science
The Hebrew University of Jerusalem
Jerusalem, Israel

1. L. AMERIO, "Su un problema di vincoli unilaterali per l'equazione non omogenea della corda vibrante," *I.A.C.* (*Istituto per le Applicazioni del Calcolo "Mauro Picone"*) *Pubbl. Ser. D.*, v. 109, 1976, pp. 3–11.

2. H. BREZIS, *Opérateurs Maximaux Monotones et Semi-Groupes de Contraction dans les Espaces de Hilbert*, North-Holland, Amsterdam, 1973.

3. C. CITRINI, "The energy theorem in the impact of a string vibrating against a point shaped obstacle," *Rend. Accad. Naz. Lincei*, v. 62, 1977, pp. 143–149.

4. C. CITRINI, "Risultati tipici sul problema della corda vibrante con ostacolo puntiforme," *I.A.C.* (*Istituto per le Applicazioni del Calcolo "Mauro Picone"*) *Pubbl. Ser. III*, v. 134, 1978, pp. 1–24.

5. M. CROUZEIX & A. MIGNOT, *Analyse Numérique des Équations Différentielles*, Masson, Paris, 1984.

6. G. LEBEAU & M. SCHATZMAN, "A wave problem in a half-space with a unilateral condition at the boundary," *J. Differential Equations*, v. 53, 1984, pp. 309–361.

7. J. L. LIONS, *Quelques Méthodes de Résolution de Problèmes aux Limites Non-Linéaires*, Dunod et Gauthier-Villars, Paris, 1969.

8. R. S. PHILLIPS, "The adjoint semi-group," *Pacific J. Math.*, v. 5, 1955, pp. 269–283.

9. R. T. ROCKAFELLAR, "Integrals which are convex functionals II," *Pacific J. Math.*, v. 39, 1971, pp. 439–469.

10. M. SCHATZMAN, "A class of nonlinear differential equations of second order in time," *Nonlinear Analysis, Theory, Methods and Applications*, v. 2, 1978, pp. 355–373.

11. M. SCHATZMAN, "Un problème hyperbolique du 2ème ordre avec contrainte unilatérale: la corde vibrante avec obstacle ponctuel," *J. Differential Equations*, v. 36, 1980, pp. 295–334.

12. M. SCHATZMAN & M. BERCOVIER, *On the Numerical Approximation of a Vibration Problem with Unilateral Constraints*, Rapport interne, Centre de Mathématiques Appliquées, Ecole Polytechnique, vol. 124, 1985.